How to define a weighting system for composite indicators

Università degli Studi di Firenze

WORKING PAPER

Filomena Maggino

The state of the art in indicators construction

To go deeper (B)

How to define weighting systems for composite indicators

Table of contents

Introduction 3

1. General concerns and principles underlying the weighting issue 5
1.1 Conditions for obtaining weights 5
1.2. Statements in obtaining weights 6
1.2.1 Equal vs. differential weighting 6
1.2.2 Weights and aggregating features: compensatory and non-compensatory techniques 7
2. Statistical approaches in obtaining weights 9
3. Statements and approaches in obtaining subjective weights 13
3.1 Obtaining subjective weights at individual or group level 13
3.2 Multi-attribute approaches 14
3.2.1 Analytic Hierarchy Processes 15
3.2.2 Conjoint analysis approach 15
3.3 Scaling approaches 16

References 19

I would like to thank Willem Saris for his invitation at the Conference on Composite Scores (ESADE, Barcelona, February 14-15, 2008) where I had the opportunity to discuss the topic and Peter Schmidt for his precious advices in preparing the preliminary version of this manuscript.
I would like to thank also Kenneth Land and Luigi Fabbris for having shared this topic with me.

This work, or part of it, has been presented in different occasions, among which:

- VII International Conference on Social Science Methodology (1-5 September 2008, Campus di Monte Sant'Angelo – Naples – Italy) – paper presented at the session "Social Indicators" (session organizer Heinz-Herbert Noll – GESIS, Mannheim, Germany).
- IV International Conference on Quality of Life Research – QOL 2008 – "Quality of Life Improvement through Social Cohesion" (Department of Statistics of Wroclaw University of Economics, 15-18 September 2008 – Wroclaw – Poland) – paper presented at the opening session.
- International Seminar on "Involving citizens/communities in measuring and fostering well-being and progress: towards new concepts and tools", Council of Europe, (Strasbourg, 27-28 November, 2008) – invited paper.
- Seminar on "*New Techniques and Technologies for Statistics (NTTS)*" – EUROSTAT (18-20 February 2009 – Brussels – Belgium) – paper.
- 62nd conference of the World Association for Public Opinion Research "Public Opinion and Survey Research in a Changing World" (Swiss Foundation for Research in Social Sciences – University of Lausanne – 11-13 September 2009 – Lausanne – Switzerland) – paper.

Introduction

The methodology aimed at constructing indicators is very often presented in terms of "technology", by asserting the need to have specialist training in order to apply the procedure in a scientific and objective way. Actually the construction procedure, even though scientifically defined, is far from being objective and aseptic.
As known, the consolidated methodology aimed at the construction of composite indicators (Nardo et al., 2005; Sharpe & Salzman, 2004) defines different stages in order to develop the indicators. Each stage requires a decision / choice (methodological or not) to be taken:

1. **choosing the analytical approach** in order to verify the underlying dimensionality of selected elementary indicators (*dimensional analysis*)
2. **choosing and obtaining weights** in order to define the importance of each elementary indicator to be aggregated (*weighting criteria*)
3. **choosing and identifying the aggregating technique** in order to synthesize the elementary indicators values into composite indicators (*aggregating-over-indicators techniques*)
4. **choosing models and conceptual approaches** in order to assess
    a. the robustness of the synthetic indicator in terms of capacity to produce correct and stable measures (*uncertainty analysis, sensitivity analysis*)
    b. the discriminant capacity of the synthetic indicator (*ascertainment of selectivity* and *identification of cut-point or cut-off values*)

Even though some decisions are strictly technical, it is quite difficult to make these decisions objective since they may involve different kind of concerns. Generally they are taken through a process accepted and shared by the scientific community.
In indicator construction, particular attention is paid to the **weighting process**. Weights aim at assigning differential *importance* weights to the indicators to be aggregated. With reference to this process, the necessity of choosing weights preferably through objective principle is always asserted (Nardo et al., 2005; Ray, 2008; Sharpe & Salzman, 2004).
However, since developing and defining weights can be always interpreted in terms of **values judgment**, the procedure should include and involve individuals' contributions in attributing importance to different domains.
Further, in certain cases, the choice and decision may be shared by a larger community. One of the ways to obtain this is that to involving individuals in the process of social indicators construction.

Subjective weights for composite indicators

In a recent work by Hagerty and Land (2007), further views were introduced about weighting in the context of composite indicators construction. In particular, creating composite indicator (describing social units at macro level) should take into account the agreement among citizens concerning the importance to be assigned to each indicator. The final composite should maximize this agreement. In their work, they provide a framework to jointly consider weights and social indicators as part of the research problem of constructing a composite indicator. This requires:

- a methodology allowing subjective weights to be collected (*subjective/individualized weighting* procedure) at individual-subjective level through subjective judgments
- a methodology allowing subjective weights to be included in indicators by assigning the weights to the corresponding indicators[1]

Subjective weights for subjective indicators

Comparison between findings concerning subjective characteristics observed at both macro (e.g. countries) and micro (cases or groups) level represents one of the more vexed issues in the field of social research and surely is among the much-discussed matters. One of the difficulties in dealing with comparison issues concerns if and how the differences might be explained, and if and how explanations could help in performing comparisons more accurately.
Among the exemplificative fields in which this topic is perceived and judged particularly sensitive (also for the implications at policy level) we can be find "measuring quality of life" where one of the goals is that to compare different levels of quality of life measured in terms of subjective well-being. According to some explanatory models, differences in well-being could be explained (Christoph and Noll, 2001) by objective characteristics, e.g. different living conditions (objective micro level) and different national structures

[1] In this context we are not discussing the procedure aimed at identifying different weight to be assigned to each case with reference to sample design. In this case the weights refers just to the level of "representativeness" (expressed in terms of proportion) that each individual of the sample has in reference to the corresponding population.

(objective macro level). Also different cultural traits and value orientations should be examined and observed at micro level and next should be properly considered in order to perform comparisons at macro level (region, country, etc.). In this perspective, the question could be how to carry on comparisons between individuals (or groups) by taking into account inter-individual (or inter-group) differences yielded by different contextual conditions, i.e. cultural traits and value orientations.

One of the possible answers may involve the definition of "subjective weights".

For example, according with the bottom-up model (*formative model of measurement*), satisfaction with life as a whole could be observed by combining satisfactions with different life ambits (family, work, income, and so on). The combination that generates the total satisfaction has to take into account the *importance* (in terms of "life value" or in terms of "expectations") that each individual assigns to each ambit. This allows scores of satisfaction to be compared by taking into account the importance assigned by individuals to each ambit.

Studies that have specifically compared weighted and unweighted scores in the field of quality of life has produced almost uniformly negative results (Andrews & Withey, 1976; Campbell et al., 1976; Cummins et al., 1994).

However, despite these negative outcomes, many researchers urge the scientific community to explore this topic by more research that specifically compares weighted and unweighted scores in particular in assessing quality of life measures (Russell et al., 2006)

- . - . - .

This work aims at showing the possible approaches in order to obtaining weights in a subjective perspective and anticipate a case study we are going to accomplish by applying and comparing all the practicable solutions. In particular, it attempts to clarify the issues to be faced in obtaining differential weights obtained. In particular, it

a) introduces the general underlying principles in obtaining weights
b) introduces the particular statements to be taken into account in obtaining subjective weights
c) identifies and analyses the approaches for obtaining:
    a. "objective weights", i.e. statistical approaches generally applied in the ambit of composite indicators construction
    b. "subjective weights", in particular:
        - Multi-Attribute approaches
        - Scaling approaches, allowing subjective data to be managed; among these, the models able (i) to handle subjective evaluations and judgments, expressed in explicit or implicit way, (ii) to obtain subjective [importance] weights at group level and at individual level, will be identified and described in the perspective of obtaining subjective weights

Pros and cons of these approaches in the perspective of subjective weighting will be discussed.

However, determining a weighting system in indicators construction is not simply a technical problem but should become part of a larger debate concerning how to construct indicators obtaining a larger *legitimacy*. Seen in this perspective, this topic can be placed in the ambit of an improvement of democratic participation to decisions ("res publica").

However, this work should take into account that the «assignment of weights that implies the knowledge of relative valuation of elementary indicators is difficult, especially when different facets of social and economic well-being are aggregated. Policy makers or government officials will most likely be unable to represent their constituents relative valuations of components in a social index. And constituents may be unable to compare such disparate aspects of well-being; indeed, this may be impossible, as peoples' preferences can be non-transitive, especially with a high number of variables. For this reason, it may be wise to abandon the notion that there exists a set of weights capable of perfectly capturing the relative contribution of variables to overall well-being» (Sharpe & Salzman, 2004).

1. General concerns and principles underlying the weighting issue

In general terms, when we suppose that not necessarily all the measured indicators (sub-score) contribute with the same importance to the measurement and evaluation of the total variable (synthetic score), a weighting system needs to be defined in order to assign a weight to each indicator, before proceeding to the indicators aggregation.
From the technical point of view, the weighting procedure consists in defining and assigning a weight to each sub-score. The weight will be used in the successive computation of the individual aggregate score; in particular, each weight is multiplied for the corresponding individual value of the sub-score.[1]
A criterion should be adopted in order to define a weighting system, when it cannot be implicitly identified. The weighting system should reproduce as accurately as possible the contribution of each sub-score to the construction of the synthetic score. In this perspective, defining a weighting system constitutes an improvement and refinement of the adopted model of measurement.
In order to proceed to the difficult choice among the different weighting approaches, the researcher needs to take into account (Nardo et al., 2005; Ray, 2008):

- the rationale and theoretical framework on which the measurement of the complex characteristics is founded and that will consequently regard the synthetic score
- the meaning and the contribution of each sub-score to the synthesis
- the quality of data and the statistical adequacy of indicators

The identification of a system of weights should

- consider in advance also some technical issues, related to the <u>conditions for obtaining weights</u> and concerning the level at which and the scale on which the weights should be determined (rescaling issue)
- make a <u>decision</u> in advance <u>on</u>:
  - the <u>proportional size</u> of the weights (equal or differential weighting)
  - the <u>aggregation technique</u> to be adopted (compensatory or non-compensatory)

1.1 Conditions for obtaining weights

The procedure for determining the weights has to take into account some <u>basic conditions</u> that can be technically formalized. As known, the general approach to composite indicators computation is the following:

$$CI_i = \sum_{j=1}^{K} x_{ij} w_{ij}$$

where

CI_i	composite indicator for case *i*
k	number of indicators to be aggregated
x_{ij}	indicator *j* to be aggregated for case *i*
w_{ij}	weight *j* to be attribute to *x* for case *i*

Each weight w_{ij} should satisfy the following basic conditions

(i) the weights are non negative numbers: $0 \le w_{ij} \le 1$

[1] An alternative to the simple multiplication of weight and score is proposed by Hsieh (2003, 2004) – and discussed by Wu (2008) – by including the sum of importance scores as a denominator. This approach can be differentiated according to ranking and rating scores when directly used as weights. Hsieh (2003) identified different computational approaches in order to connect the weight with the score to be weighted. In particular, he proposed seven different weighting mechanisms of relative importance, three using discrete importance rating and four using ranking scores. The sum of weighted scores is divided with differently adequate denominators to obtained equivalent scales for the purpose of easy and intuitive comparison.
Since in collecting subjective data to be directly used as weights, both scores can be adopted, rating and ranking scores data should be carefully assumed by considering that they can reflect different meanings in terms of weight and require different computational approaches, producing different results.

(ii) the weights for each case *i* add up to unity: $\sum_{j=1}^{K} w_{ij} = 1$

(iii) the weights may require to be rescaled in order to have an identical range

(iv) the weights are relating in some way to the corresponding score (as we will see, this condition may require a decision to be taken)

Rescaling weights

Following their computation, weights may require to be rescaled. Re-scaling
- normalises weights to have an identical range (0; 1)
- could distort the transformed indicator in presence of extreme values/or outliers
- could widen the range of indicators lying within a small interval increasing the effect on the weights.

The procedure can be performed as follows:

$$rw_{ij} = \frac{w_{ij}}{\max(w_j)} \quad \text{or} \quad rw_{ij} = \frac{w_{ij} - \min(w_j)}{\max(w_j) - \min(w_j)}$$

where

rw_{ij} rescaled value of the weight with reference to the object *j* for the respondent *i*

w_{ij} value of the weight with reference to the object *j* for the respondent *i*

The researcher has to carefully evaluate and make formally explicit not only the methodology to be adopted but also the results that would have been obtained with other methodologies, also reasonably applicable.

1.2 Statements in obtaining weights

1.2.1 Equal vs. differential weighting

The first decision that needs to be made and that will be strongly influence the final results is between *Equal Weighting* (*EW*) and *Different Weighting* (*DW*).

Equal weighting represents the preferred procedure, adopted in most of the applications. This happens mainly when:
- the theoretical structure attributes to each indicator the same adequacy in defining the variable to be measured
- the theoretical structure does not allow hypotheses to be consistently derived on differential weightings
- the statistical and empirical knowledge is not adequate for defining weights
- the correct adoption and application of alternative procedures do not find any agreement

Although equal weighting, which does not necessarily imply unitary weighting, is certainly an explicit weighting scheme, the a priori decision to adopt the technique of equal weighting for methodological purposes makes the choice of weights apparently less subjective. A motivation for this approach is that it is objective in the sense that if adopted as a common technique of weighting, the subjective component would lie exclusively in the choice of indicators. There is an advantage of this approach: namely, that a debate over the inclusion of elementary indicators, that is, which indicators are important, can be conducted on a more basic level than a discussion that focuses on the choice of numerical weights (Sharpe & Salzman, 2004).

Another strength of this approach is that if indicators are chosen as indicators for something that cannot be perfectly quantified, from the perspective of the social indicator constructor, the indicators chosen as variables for a category of measurement should form a collection of multidimensional indicators that is a sampling of indicators that may represent the category. Since the elementary indicators are indicators and not measurements in themselves, it is more consistent to treat them as statistical objects that are not subject to further subjective numerical interpretation. As we will discuss below, it does not always make sense to apply any differential weighting to social measures due to the complex nature of social and economic phenomena. As a result, the case for uniformly aggregated variables, that is, a priori equal weights, is strengthened (Sharpe & Salzman, 2004).

Equal weighting procedure can be doubtful when:
- the definition of the variable requires different components specified by different numbers of indicators; in this case, adopting equal weighting corresponds to assigning higher weights to the components showing higher numbers of elementary indicators; in these cases, the synthetic variable will have an unbalanced structure;

- the existence of indicators measuring the same component (high correlations between elementary indicators): the result corresponds to that obtained when higher weights are assigned to indicators showing higher correlation (*double weighted* o *double counting*).

Differential weighting does not necessarily correspond to the identification of different weights but rather to the selection of the most appropriate approach in order to identify the weights among the identified ones (Nardo et al., 2005).

Assigning differential weights can be just as doubtful, especially when the decision is not supported by:

- theoretical reflections that endow a meaning on each indicator or consider its impact on the synthesis,
- methodological concerns that helps to identify the proper techniques, consistently with the theoretical structure.

In any case, we have to consider that a whole set of weights able to express in a perfect way the actual and relative contribution of each indicator to the measurement does not exist.

Independently from the approach adopted in order to define them, the weights can be kept constant or can be changed according to particular considerations concerning each application. In both cases, the researcher needs to rationalize the choice. The former approach can be adopted when the aim is to analyse the evolution of the examined ambit. The latter can be adopted when the aim concerns – for example – the definition of particular priorities.

Bobko et al. (2007) made a interesting review of the relevant literature across multiple disciplines and multiple decades on differential and unit weights. Their literature review indicates that unit weights have substantial predictive validity when compared with regression weights, but there is a lack of data on how other differential weighting strategies (e.g., weights generated by subject matter experts) compare to unit weights. Moreover, they provide a primary and a meta-analytic study by which they show how in their applications data and findings indicate that unit weights can be a highly appropriate approach for weighting under many circumstances.

In subjective measurement, the effectiveness of weighted scores should be questioned with reference to

(i) the theoretical issue of whether importance and satisfaction are distinct constructs,

(ii) the psychometric properties of importance ratings (particularly, internal consistency and test-retest reliability), and

(iii) the criteria used in assessing weighted scores.

All these topics need more attention and care from the researchers (Russell & al., 2006)

1.2.2 Weights and aggregating features: compensatory and non-compensatory techniques

In order to avoid incoherencies between the theoretical meaning of weights and the way these weights are actually used, a consistent aggregating technique has to be chosen (Nardo et al., 2005).

Let us formally represent the issue:

In particular, the choice of the weighting system must consider the compensability among the elementary indicators inside the synthetic score. In particular, this is allowed by the technique that will be used in aggregating the sub-scores.

An aggregating technique is <u>*compensatory*</u> when it allows low values in some sub-scores to be compensated by high values in other sub-scores. In the typical aggregating table (see below), we can observe all the possible synthetic scores, obtainable by aggregating two sub-scores (A and B) using simple addition (additive approach):

		B		
		1	***2***	***3***
A	***4***	5	6	7
	3	4	5	6
	2	3	4	5
	1	2	3	4

Some of the obtained synthetic values, even if completely identical, are obtained through different sub-scores. This means that the obtained aggregated values do not allow us to return to the original unit's profile. In other words, two units relating with different realities turn out to be identical and not distinguishable from each other.

By using the same data reported in previous table, all the possible synthetic values can be observed, obtainable by aggregating two indicators (A and B) using a multiplicative technique (geometrical approach):

		B		
		1	2	3
A	4	4	8	12
	3	3	6	9
	2	2	4	6
	1	1	2	3

The table suggests that also multiplicative technique is compensatory as well, especially with reference to low scores.
Generally, in order to make multiplicative functions more manageable, the values of involved indicators are logarithmically transformed (summing up logarithm values corresponds to multiplying the original values). However, this procedure has to be followed with caution since it can also produce problems of interpretation.
If compensability is admitted, a unit showing a low value for one indicator will need higher values on the others in order to obtain a higher synthetic value.

2. Statistical approaches in obtaining weights

One of the ambit in which the issue of obtaining differential *importance* weights found consolidated applications is that of constructing composite indicators. As previously said, in this ambit, it is always asserted how the choice of weights would be preferably derived from objective principle (Ray, 2008). In this perspective, the statistical methods are traditionally considered and preferred (Nardo et al., 2005; Ray, 2008; Sharpe & Salzman, 2004), above all (i) *Correlation Analysis* (CA), (ii) *Principal Component Analysis* (PCA), (iii) *Data Envelopment Analysis* (DEA).
The adoption of statistical methods in weighting components of social indices has to be considered carefully since, by removing any control over the weighting procedure from the analysts, it gives a false appearance of mathematical objectivity that is actually difficult to achieve in social measurement (Sharpe & Salzman, 2004).

Correlation Analysis

As previously said, assigning equal weights to elementary indicators that are highly correlated can introduce the *double counting* effect. By contrast, the correlation values can be considered as assigning a weight to each elementary indicators. This can be done by averaging the correlation values registered between all the selected elementary indicators. In particular, this weight can be inversely proportional to the correlation level; this approach allows less importance to be assigned to indicators that are highly correlated to the others. The application of this approach leads to the definition of a limit value that allows a high correlation that could be considered as a sign of *double counting* to be identified. The limit can not be defined at a statistical level because there is no statistical rule on this matter; in any case, such decision can not be made on a statistical base but in the ambit of the adopted conceptual framework.

Principal Component Analysis (PCA)

The goal of principal component analysis is essentially to uncover variations in a data set. Principal component analysis can be used to describe the variation of a data set using a number of scores that is smaller than the number of the original elementary indicators.
This approach is particularly useful in the case of multidimensional latent variables since its algorithms enable the weight (*component score*) to be assigned to each elementary indicator to be determined, subsequently the identification of the components explaining the greatest portion of total variance.
The weights of the components in the first dimension, which is called first principal component, are assigned to maximize the variation in the linear combination of original variables, or (equivalently) to maximize the sum of the squared correlations[1] of the principal component with the original variable. Another way to think about this is that the first principal component is represented by the line in the original space of variables that minimizes the sum of the squared distances between it and the original data points.
The weights allow one synthetic indicator for each component to be calculated (Dunteman, 1983). Component scores measure the independent and not-correlated contribution of each elementary indicator in defining each component[2]. It can be calculated by removing the part of the contribution explained by its correlation with the other elementary indicators. This is because values of component scores are usually lower than the respective component loading. When the identified components perfectly reflect the existing dimensional structure (previously tested through factor analysis) and each elementary indicator has only one significant component score, the resulting synthetic indicators will be consistent and independent from each other. The adoption of this approach has to consider that the meaning of the weights (component scores) is exclusive statistical.

Data Envelopment Analysis (DEA)

This nonparametric method belongs to the group of those approaches developed in operation researches and economics and that are aimed at studying and evaluating the efficiency/inefficiency of production processes through the definition of production frontiers.
The objective of *DEA* is to estimate the *efficiency frontier* that would subsequently be used as a benchmark in order to measure and evaluate the relative performance of observed units (said *Decision Making Unit*, DMA). This evaluation is made in terms of the distance of each DMA from the efficiency frontier. The group of weights are derived from the comparisons carried out.
In this perspective, DEA can represent a valid approach in order to identify weights to be assigned to elementary indicators, with particular reference to those indicators related to concepts like “capacity”. This

[1] The correlation matrix is the covariance matrix of variables, which are scaled in order to have unit variance.
[2] The weights cannot be represented by the component loadings since these indicate only the importance and the validity of the elementary indicators in defining the general concept (latent variable) and its components.

approach, formerly developed by Charnes, Cooper and Rhodes (1978), measures the efficiency of multiple Decision Making Units (DMUs) by a Linear Programming methodology when the production process presents a structure of multiple inputs and outputs.
The benefits of using DEA are due to the following factors:

- there is no need to explicitly specify a mathematical form describing the production function and the performance model,
- it is useful in uncovering relationships that remain hidden for other methodologies,
- it is capable to handle many elementary indicators at the same time,
- it is possible to use it with any kind of input-output measurement,
- the sources of inefficiency can be identified, analysed and quantified for every evaluated unit (in other words, it is possible to identify the elementary indicator showing the worse performance).

The distance of each unit with respect to the benchmark is determined by the location of the unit and its relative position with respect to the frontier. The performance indicator is the ratio of the distance between the origin and the actual observed point and the projected point in the frontier. The best performing units will have a performance score of 1, while the least performing less than one. The set of weights for each unit depends on its position with respect to the frontier, while the benchmark corresponds to the ideal point with the same group of elementary indicators (Nardo et al., 2005).
The procedure called *Benefit-Of-the-Doubt* (BOD) is considered a special case of DEA. This procedure allows a different priority to be emphasized or defined for each observed case according to those aspects that turned out to be good performances; in some sense this requires the individual identification of a strategic- or priority-objective (*target*) instead of identifying an *efficiency frontier* (Nardo et al., 2005). If no restriction is set to the definition of the best individual performance, the optimizing procedure could lead to the definition of null weights. For this reason, the actual use of this approach requires the identification of restrictions to the individual targets and consequently to the weights.

2. Statistical approaches in obtaining weights

	Correlation Analysis	Principal Component Analysis	Data Envelopment Analysis
Assumptions	Assigning equal weights to elementary indicators that are highly correlated.	The group of elementary indicators define a multidimensional variables.	The elementary indicators define particular dimensions, like capacity.
Goals	Less importance to be assigned to indicators that are highly correlated to the others (*double counting* effect)	Definition of a set of weights each dimension/component defining the latent variable for each elementary indicator – components scores. Weights allow one synthetic indicator for each component to be calculated. The resulting synthetic indicators will be independent from each other.	Identifying weights to be assigned to elementary indicators, with particular reference to concepts like "capacity". *DEA* estimates the *efficiency frontier* that can be used as a benchmark in order to measure and evaluate the relative performance of observed units.
Methods	Since a high correlation is considered evidence of *double counting*, the procedure requires • averaging the correlation values registered between all the selected elementary indicators • defining the weight (in inverse proportion as regards the correlation level)	• identification of the components explaining the greatest portion of total variance • for each elementary indicator, calculation of component weight by removing the part of the elementary indicator's contribution explained by its correlation with the other elementary indicators.	The set of weights for each case depends on its position defined in terms of its distance from • the efficiency frontier (corresponding to the best registered performance) or • the benchmark (corresponding to an ideal point) or different priority defined according to those aspects that turned out to be good performances (this requires the individual identification of a strategic- or priority-objective).
Benefits			It allows us to • avoid an explicit specification of a mathematical form describing the production function and the performance model • uncover relationships that remain hidden for other methodologies • handle many elementary indicators at the same time. • use any kind of input-output measurement • identify, analyze, and quantified the sources of inefficiency, for every evaluated unit (in other words, it allows us to identify the elementary indicator showing the worse performance)
Limits	The **limit value** can not be defined at a statistical level because there is no statistical rule on this matter.	The **weights' meaning** (component scores) is exclusive statistical.	The approach is **not always applicable**, especially in presence of many elementary indicators.

<u>*Unobserved Components Models*</u> (UCM).
This approach retrieves the reflective rationale (elementary indicators depend on an unobserved variable plus an error term). In this perspective, estimating the unknown component sheds some light on the relationship between the synthetic indicator and its components (elementary indicators). The obtained weight will be set in order to minimize the error in the synthetic indicators by applying a *maximum likelihood* approach. According to this approach, the observed score corresponds to

$$I_{iq} = \alpha_q + \beta_q \left(ph_c + e_{iq} \right)$$

where

ph_c	unobserved component
$q = 1,...,Q_c$	group of elementary indicators, each indicator measures an aspects of ph_c
I_{iq}	score observed by the *i* unit for *q*
e_{iq}	error term concerning the score observed by the *i* unit
α_q e β_q	unknown parameters concerning the location of ph_c on I_{iq}

In order to apply UCM, the reflective model's assumptions need to be met:
- the error term represents an independent variable,
- the error term includes two uncertainty sources: measurement error and imperfect relationships between elementary indicators and synthetic indicators,
- the unknown synthetic indicator (ph_c) represents a random variable with mean=0 and variance=1,
- elementary indicators should be properly rescaled (from zero to one),
- ph_c and e_{iq} are normally distributed.

By mean of the conditional distribution of the unobserved component, the weights are obtained as follows:

$$w_{iq} = \frac{\sigma_q^{-2}}{1 + \sum_{q=1}^{Q_i} \sigma_q^{-2}}$$

where w_{iq} represents the weight for the case *i* and the indicator *q*.

Actually, w_{iq} represents
- a decreasing function of the variance of indicator *q* (the larger is the variance of the indicator, the smaller is its precision and the smaller is the weight assigned to the indicator),
- an increasing function of the variance of the other indicators.

The equal number of elementary indicators for all the cases ensures that comparability among cases is assured.

3. Statements and approaches in obtaining subjective weights

In order to identify a subjective weighting system, a **model** should be chosen by consideringthe criterion of importance or preference to be adoptedthe level at which weights are determined (*individual* or *group* weights)

- the techniques allowing subjective evaluations and judgments to be expressed by subjects in a directly or indirectly way
- the approach allowing a subjective importance/preference continuum to be constructed in order to transform evaluations and judgments into data analyzable and interpretable in terms of importance/preference weights.

3.1 Obtaining subjective weights at individual or group level

In order to determine subjective weights,

- data should be collected at individual level
- weights can be defined at
  - individual level: individual data will be used in order to construct weights that could be different for each subject,
  - group level: individual data will be used in order to construct different weights for different group of individuals.[1]

The issue can be formally represented as follows:

[1] In both cases, the general basic conditions described above are equally valid in obtaining subjective weights.

Subjective weighting at individual level		
Let us define		
For *K* objects, the set of weights for the individual *i* must satisfy the following basic conditions:		
(i)	$0 \leq w_{ij} \leq 1$	the weights are non negative numbers:
(ii)	$\sum_{j=1}^{K} w_{ij} = 1$	the weights add up to unity
(iii)	$z_{ij} = x_{ij} * w_{ij}$	the weighted score is obtained by relating *x* to *w* in some way:

Subjective weighting at group level		
Let us define		
For *K* objects, the set of weights for the group *c* must satisfy the following basic conditions:		
(i)	$0 \leq w_{cj} \leq 1$	the weights are non negative numbers:
(ii)	$\sum_{j=1}^{K} w_{cj} = 1$	the numbers add up to unity:
(iii)	$z_{icj} = x_{icj} * w_{cj}$	the weighted score is obtained by relating *x* to *w* in some way:

Subjective weighting at individual level	
Let us define	
X	a matrix with *N* rows (*i*=1 … *N*, individuals) and *K* columns (*j*=1 … *K* object variables) in which x_{ij} score that individual *i* assigned to *j* object (e.g. satisfaction for family)
W	a matrix with *N* rows (*i*=1 … *N*, individuals) and *K* columns (*j*=1 … *K* object variables) in which w_{ij} importance that individual *i* assigned to *j* object (e.g. importance of family)
Z	a new matrix with *N* rows (*i*=1 … *N*, individuals) and *K* columns (*j*=1 … *K* weighted object variables) in which z_{ij} weighted score for individual *i* concerning *j* object

Subjective weighting at group level	
Let us define	
X	a matrix with *N* rows (for *i*=1 … *N*, individuals) and *K* columns (for *j*=1 … *K* object variables) in which x_{cij} score that individual *i* belonging to the *c* group assigned to *j* object (e.g. satisfaction for family) The group can be predefined or can be determined through clustering methods
W	a matrix with *G* rows (for *c*=1 … *G*, groups) and *K* columns (for *j*=1 … *K* object variables) in which w_{cj} importance that group *c* assigned to *j* object (e.g. importance of family)
Z	a new matrix with *N* rows (for *i*=1 … *N*, individuals) and *K* columns (for *j*=1 … *K* weighted object variables) z_{ij} weighted score for individual *i* concerning *j* object

The aim is
a. to determine the values of the W matrix (in the two versions, weights for individual and weights for groups)
b. to determine the interpretable values in Z matrix
c. to sum up the K weighted scores in a unique individual synthetic score.

In the following paragraphs, methods supporting the two perspectives, individual and group weighting, will be discussed.

3.2 Multi-attribute approaches

In order to define importance of a group of elements (elementary indicators) to be identified at subjective level and consequently to identify subjective weights methods are required able to manage a certain number of combined comparisons. These comparisons can be managed by applying methods aimed at making decision among different available alternatives. These methods are encompassed among *Multi-Attribute Models* Usually. Weights obtained through these methods are considered more stable than those produced by direct evaluations. Among these models we can distinguish:

a. ***Multi-Attribute Decision Making*** (*MADM*): it represents a branch of the wider field of *Multiple Criteria Decision Making* (MCDM) and refers to making preference decisions (e.g., evaluation, prioritization, selection) over available alternatives that are characterized by multiple conflicting attributes (Yoon, 1995). **Analytic Hierarchy Process (AHP)** (*pairwise comparison of attributes*) represents one of the techniques used in this ambit.
b. ***Multi-Attribute Compositional Models***: these models are based upon a statistical de-compositional approach through which it is possible to manage subjective comparisons of attributes on different levels. Its goal is to determine which combination is preferred by the subject. Among these model, **Conjoint Analysis** (CA) is the most known. While AHP approach derives the “importance” of an alternative by summing up the scores of the elementary indicators, CA approach proceeds in the opposite direction, that is by disaggregating the preferences, expressed by the subject in combination (Edward & Newman, 1982; Yoon, 1995).

3.2.1 Analytic Hierarchy Processes

***Analytic Hierarchy Processes* (AHP)** represents a structured technique for dealing with complex decision. AHP provides a comprehensive and rational framework for structuring a problem, for representing and quantifying its elements, for relating those elements to overall goals, and for evaluating alternative solutions. This approach proceeds by decomposing the problem related to the decision in hierarchical terms (sub-problems that be more easily and independently comprehended and analyzed), aspects that are both qualitative and quantitative can be to embodied in the evaluating process. Many solutions, provided by pros and cons, can be analyzed and compared.
AHP is considered a compensative methodology since the identified alternatives can turn out to be efficient with regard to one or more objectives which counterbalance their performances.
AHP is based upon three basic principles:

- Interacting and interrelated attributes (objects) are not allowed (independency of criteria); the preferences that can be expressed regarding the different alternatives depend upon separate attributes which can be separately sustained and to which numerical scores can be assigned.
- Attributes can be hierarchically organized and the score for each level of the hierarchy can be calculated by summing up the weighted scores of the lower levels; this assumption does not admit attributes presenting a threshold.
- Scores can be calculated for each level from paired comparisons data; this can be performed only if the number of items is quite low (with 4 alternatives, the comparisons are 6 (=4*3/2) while with 20 alternatives, the comparisons are 190).

The AHP presents some characteristics that can lead to identification of various types of errors in decision:

- possible different hierarchies can be identified in applying to identical problems
- possible major changes in results if the hierarchy is changed in minor ways
- absence of statistical theory to underlie the process
- use of arbitrary scales: AHP is mainly based on pairwise comparisons where the relative importance of different attributes are given a value on a scale of 1 to 9 or the inverse (1/9th to 1) with all the problems of arbitrariness that this implies. A good approach could be the identification and the proposal of different alternative scales
- possible inconsistent judgments: AHP, like many procedures based on pairwise comparisons, can produce "rank reversal" outcomes producing inconsistent results (a respondent might have said X is preferred to Y, Y to Z but Z is preferred to X). However, since any pairwise comparison system has rank-reversal solutions even when the pair preferences are consistent, some analytic corrections were defined in order to deal with this problems
- risk to induce ordering even when no order actually exists. This problem can reveal the lack of clear definition of the conceptual framework.

Applicability of AHP in order to obtain subjective weights

In our perspective (obtaining subjective weights), the possibility to identify different hierarchies when applied to identical problems can turn out to be some kind of advantage, represented mainly by the possibility to obtaining subjective weights at individual level by a quite straightforward approach. However, the need to construct a hierarchy with many nodes might make this approach non-applicable in the context we are dealing with (large surveys).

3.2.2 Conjoint analysis approach

Applicability of conjoint model in order to obtain subjective weights

The approach, presented in a different context,[2] can be performed at both individual and group level. In particular, the choices expressed by a group of subjects can be combined in order to represent a “competitive” ambient.
This approach is considered *compensatory* and consequently requires a careful evaluation of its applicability. The estimated part-worths allow the range of importance for each factor to be determined. By dividing each factor’s range by the sum of all range values we can obtain the proportion, interpretable in terms of importance of each factor in the respondent’s choice. The polarity is consistent to the response scale submitted to the respondents and is considered inside the analytical procedure. The approach

- allows obtained proportions to be assigned to objects in terms of weights
- does not require the rescaling procedure to be applied

[2] With reference to this issue, see the first contribution in “to go deeper” section: *Methodological aspects and technical approaches in measuring subjective well-being*.

- does not allow a continuum of importance to be obtained
- meets the requirement of the sum of weights (sum of the obtained proportions is equal one)
- can be applied for obtaining subjective weights at both individual and group level.

However, the approach should be applied with great caution since the obtained weights strongly depend upon the definition of the levels for each factor.

3.3 Scaling approaches

As known, the traditional approaches that enable to deal with subjective evaluations and judgments are the "**scaling models**". Let recall the features that can describe and characterize each scaling model (McIver & Carmines, 1979; Maggino, 2007):

- **Dimensionality**, concerning the variable to which the combined individual score/s will be referred.
- **Nature of data**, referring the Coombs' classification (single stimulus, stimulus comparison, similarities, preferential choice).
- **Scaling technique**, referring to the comparative and non-comparative approaches.
- **Criterion for testing the model**, referring to checking model data fitting.
- **Standard of measurement**, concerning the treatment of the multiple measures and the assignment of the synthetic value (the final score can be assigned to individuals or to stimulus).
- **Contribution of each multiple measures to the measurement**:

The following table (Maggino, 2007) summarizes the characteristics of the well-known scaling models and allows us to identify those that better can help us in pursuing our goal, the identification of subjective weights.

In particular, since we are looking for a "subjective weight" that is able to give back the idea of "subjective importance" attributed to each element (item) in comparison with the other elements composing the set, we have to select those models that utilize data

- whose nature is comparative or preferential (marked in yellow in the previous table)
- produced by a comparative scaling technique (marked in pink in the previous table).

At this point, the models that can be selected are:

- *Thurstone model* (*differential scale*) and *Q* methodology[3], comprised among the cumulative approaches,
- *unfolding model* and *conjoint model*, comprised among the "perceptual mapping" approaches.[4]

Since we need also to identify a procedure that can be applied in a survey context without particular efforts, the *Q* methodology will be excluded by our consideration.

In our perspective, these models can be distinguished with reference to the possibility to define subjective weights at individual level or at group level (last column of the previous table), in particular:

- individual weighting: *conjoint model* (again)
- group weighting: *Thurstone model* (*differential scale*), *unfolding model*.

[3] The well-known method called *Budget Allocation* (BAL) can be assimilated to Q methodology: each respondent is asked to distribute a certain budget – constituted by an *X* scores – among the objects, by assigning higher scores to those objects that he/she considers more important. In some cases, the procedure can be extended in order to achieve weights through agreement among respondents (group-weights). This approach turns out to be practicable in case of low number of objectives (max 10-12) in order to save respondents a difficult and complicated task. The approach is generally applied with a small group of subjects (experts, policy makers or public opinion polls).

[4] *Perceptual mapping* represents an approach that attempts to visually display the perceptions of individuals. Typically, each element's (item) position is displayed relative to their competition. Perceptual maps can have any number of dimensions but the most common is two-dimension.

				Scaling model's Characteristics				
				Dimensionality	**Nature of data**	**Scaling technique**	**Criterion for testing the model**	**Standard of measurement: final (synthetic) score assigned to**
Scaling models	**Additive**	Uni-dimensional		Uni	Single-stimulus	Not-comparative	Internal consistency	Cases
		Multidimensional		Multi	Single-stimulus	Not-comparative	Dimensionality of the items	Cases
	Cumulative	Thurstone model (differential scale)		Uni	Stimulus comparison	Comparative (pair comparison or rank-order)	Metrics between items	Items
		Q methodology		Uni	Stimulus comparison	Comparative (rank-order or comparative rating)		Items
		Deterministic	Guttman	Uni	Single-stimulus	Not-comparative	Scalogram analysis: reproducibility, scalability and ability to predict	Cases and items
			Multidimensional Scalogram Analysis (MSA)	Bi			Regionality and contiguity	Cases and items
			Partial Ordered Scalogram Analysis (POSA)	Bi			Correct representation	Cases and items
		Probabilistic	Monotone (one or more parameters)		Single-stimulus	Not-comparative	• parameters estimation (maximum likelihood) • goodness of fit (misfit and residuals analysis)	Cases and items (without condensation)
	Perceptual Mapping	Multidimensional scaling		Multi	Similarities	Comparative (pair comparison)	Goodness of fit of distances to proximities (stress, alienation)	Items
		Unfolding		Uni & Multi	Preferential choice	Comparative	Goodness of fit of distances to ordinal preferences	Cases and items
	Conjoint model			Multi	Preferential choice	Comparative (rank-order)	Goodness of fit of the model (part-worth) to the ranking	Items at individual level

Cumulative approach

The approach based upon the logic that can be defined as "cumulative" has the goal to "create" a continuum on which the elements (items) concerning a certain characteristic are positioned. In order to pursue this goal, the judgments expressed by a group of individuals are employed. The judgments can be expressed using the "paired comparison" scaling technique or the "rank ordering".

Applicability of cumulative model in order to obtain subjective weights

The cumulative approach
- allows a continuum of importance to be obtained
- requires the continuum to be interpreted in terms of polarity
- allows the objects to be positioned on this continuum according to a quantitative value interpretable in terms of weights
- produces weights that should be rescaled in order to meet the weights' conditions presented above.

Unfolding approach

Applicability of unfolding model in order to obtain subjective weights

The unfolding approach, presented in a different context,[5]
- allows a continuum of importance to be obtained
- requires the continuum to be interpreted in terms of polarity
- allows the objects to be positioned on this continuum according to a quantitative value interpretable in terms of weights
- produces weights that should be rescaled in order to meet the weights' conditions presented above.

[5] With reference to this issue, see the first contribution in "to go deeper" section: *Methodological aspects and technical approaches in measuring subjective well-being*.

References

Andrews, F.M. and Withey, S.B. (1976) *Social Indicators of Well-being: Americans' Perceptions of Well-being*, Plenum Press, New York.

Bobko P., P.L. Roth, M.A. Buster (2007) "The Usefulness of Unit Weights in Creating Composite Scores. A Literature Review, Application to Content Validity, and Meta-Analysis", Organizational Research Methods, vol. 10, n. 4, 689-709.

Campbell, A., Converse, P.E. and Rodgers, W.L. (1976) *The Quality of American Life: Perceptions, Evaluations, and Satisfactions*, Russell Sage Foundation, New York.

Charnes, A., W.W. Cooper, and E. Rhodes (1978) "Measuring the Efficiency of Decision Making Units", *European Journal of Operational Research*, 2(6), pp. 429-444.

Christoph B. and H.-H. Noll (2001) "Subjective Well-Being in the European Union during the 1990ies" in J. Vogel (ed.) *European Welfare Production: Institutional Configuration and Distributional Outcome*, Dordrecht, Kluwer.

Cummins, R.A., McCabe, M.P., Romeo, Y. and Gullone, E. (1994) "The Comprehensive Quality of Life Scale (ComQol): Instrument development and psychometric evaluation on college staff and students", *Educational and Psychological Measurement*, n. 54, pp. 372–382.

Dunteman G.H. (1983) *Principal Component Analysis*, Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-069, Newbury Park, CA:Sage.

Edward W., Newman J.R. (1982) *Multiattribute Evalutation*, Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-026, Newbury Park, CA:Sage.

Hagerty, M.R., Land, K.C. (2007) "Constructing Summary Indices of Quality of Life: A Model for the Effect of Heterogeneous Importance Weights", *Sociological Methods and Research*, Vol. 35, n. 4, 455-496.

Hsieh, C.M. (2003) "Counting importance: The case of life satisfaction and relative domain importance", *Social Indicators Research*, n. 61, pp. 227–240.

Hsieh, C.M. (2004) "To weight or not to weight: The role of domain importance in quality of life measurement", *Social Indicators Research*, n. 68, pp. 163–174.

Maggino F. (2007) *La rilevazione e l'analisi statistica del dato soggettivo*, Firenze University Press, Firenze.

McIver J.P., Carmines E.G. (1979) *Unidimensional Scaling*, Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-024, Newbury Park, CA:Sage.

Nardo M., M. Saisana, A. Saltelli and S. Tarantola (EC/JRC), A. Hoffman and E. Giovannini (OECD) (2005) *Handbook on Constructing Composite Indicators: Methodology and Userguide*, OECD, Statistics Working Paper.Ray A.K. (2008) "Measurement of social development: an international comparison", *Social Indicators Research*, vol. 86, n.1, 1-46.

Russell L.B., A.M. Hubley, A. Palepu and B.D. Zumbo (2006) "Does Weighting Capture What's Important? Revisiting Subjective Importance Weighting with a Quality of Life Measure", *Social Indicators Research*, vol. 75, n.1.

Sharpe A., J. Salzman (2004) *Methodological Choices Encountered in the Construction of Composite Indices of Economic and Social Well-Being*, Center for the Study of Living Standards, Ottawa, Canada.

Yoon K.P., Hwang Ching-Lai (1995) *Multiple Attribute Decision Making*, Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-104, Thousand Oaks, CA:Sage.

Wu, C. H. (2008) "Can We Weight Satisfaction Score with Importance Ranks Across Life Domains?", *Social Indicators Research*, 86, 469–480.